



Challenges in bioinformatics and computational biology

Dubitzky, W. (2011). Challenges in bioinformatics and computational biology.

[Link to publication record in Ulster University Research Portal](#)

Publication Status:

Published (in print/issue): 01/01/2011

Document Version

Author Accepted version

General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.

Editorial

Bioinformatics and computational biology refer to an interdisciplinary field which is concerned with the development and application of techniques from computer science, mathematics and statistics to address biological problems. One could argue that bioinformatics concentrates on techniques facilitating the acquisition, storage, organization, archiving, analysis and visualization of biological and medical data. Computational biology, on the other hand, focuses on theoretical methods, mathematical modeling and computational simulation techniques to study biological systems and processes.

My view is that both areas—bioinformatics and computational biology—are merging into a single but broader discipline and that the distinction between the two is becoming increasingly blurred. This is also evidenced by the growing number of scientific journals and conferences that carry the terms bioinformatics as well as computational biology in their title. One could even argue that bioinformatics and computational biology and the biological sciences are converging to an even broader discipline called systems biology. Recent developments in bioinformatics and computational biology have a tendency to attempt a more holistic understanding of complex biological processes and systems. This is partly pushed by advances in experimental technologies and pulled by a constantly increasing capacity and sophistication of information and communication technologies. The merging of the two disciplines and the trend towards a more comprehensive view of living systems is reflected by the contributions in this issue of *Briefings in Bioinformatics*.

One important consequence of the ever-increasing power of experimental techniques and simulation systems (generating data from *in silico* experiments) is the growing volume of experimental data and derived data obtained from integration with other data, annotation, etc. Furthermore, analysis and interpretation of experimental and derived data yields considerable volumes of information and knowledge in the form of scientific literature, ontologies, models and so on. While more data, information and knowledge opens up great opportunities for ‘seeing’ an increasingly ‘bigger

picture’, it also poses considerable challenges to data handling, data integration, data analysis, modeling and simulation, and knowledge management [1]. The articles in this issue present an overview of the state of the art of bioinformatics and computational biology methods and technologies that tackle challenges arising from the increasing data and knowledge complexity of modern biology. Below I briefly outline the motivation of this issue’s articles.

Kei-Hoi Cheung *et al.* investigate the status of data integration in neuroscience in the light of the rapidly increasing number and content of neuroscience databases. Their analysis covers some of the major neuroscience databases and revolves around several approaches to the problem of registering, discovering and integrating neuroscience databases.

Mihai Pop reviews recent developments of algorithmic approaches to genome assembly, taking into account next-generation sequencing technologies and new challenges such as metagenomics which relies on large-scale sequencing of entire microbial communities instead of isolated genomes.

Francisco Azuaje *et al.* have analyzed contemporary computational biology approaches to cardiovascular biomarker discovery based on ‘omic’ information. The investigated methods and technologies include predictive modeling and integration of different types of data and knowledge for screening, diagnostics and prognostics applications.

Jeffrey Skolnick and Michal Brylinski review the status of sequence- and structure-based approaches to protein function inference and ligand screening that can provide functional insights for a significant fraction of the ~50% of open reading frames with unassigned function. The context of this review is the shifting of emphasis from the study of individual molecules to a large-scale examination of all genes and gene products within a single organism or across multiple organisms. Furthermore, due to complexity arising from large evolutionary distances in ~50% of the open reading frames in a given proteome, it is suggested that sequence-based methods appear to have reached a limit in accurately predicting their function. Since protein structure is more conserved than protein *sequence*, protein *structure* could

play an essential role in annotating genomes. This view is one of the motivations for this review.

Erick Antezana *et al.* look beyond the growing mountains of experimental data and anticipate the rapid approach of another huge challenge for the life sciences—the surging oceans of biological knowledge. Biological knowledge captured by literature, ontologies, models, encyclopedias, expert systems, etc. is threatening to turn into a problem rather than an asset. Their article reviews the benefits, trends, current possibilities, and the potential of several initiatives and information technologies to organize biological knowledge sources into a readily exploitable ‘resourceome’.

Wei-Po Lee and Wen-Shyong Tzou’s article is concerned with computational techniques facilitating the inference of biological networks of varying levels of accuracy and complexity. Guided by the scientists’ motivation to generate testable hypotheses from the inferred networks, this review focuses on methods for predicting gene-regulatory networks in mammalian cells. The authors also show how the power of different databases can be used to identify modules and sub-networks, thereby reducing complexity and facilitating the generation of testable hypotheses.

Attila Csikász-Nagy reviews contemporary computational systems biology approaches aimed at an improved and more comprehensive understanding of the cell cycle. While detailed and validated mathematical and computational representations of the cell cycle exist, there are still considerable gaps in our knowledge of the cell cycle. For instance, we need to understand how the core cell cycle machinery is controlled by internal and external signals in simple and complex organisms.

Karthik Raman and Nagasuma Chandra analyze the status of research and development on the methods and tools of flux balance analysis to model and understand biological systems. Unlike

mechanistic simulations, flux balance analysis is based on the stoichiometric matrix (representing complex reaction networks) and a biologically relevant objective function to identify optimal reaction flux distributions within a system.

Dawn C Walker and Jennifer Southgate take a look at contemporary work on agent-based modeling and simulation in biology with a particular focus on multi-scale phenomena, ‘middle-out’ models and a view that positions the cell (and not the genome) as fundamental abstraction unit.

Niall Palfreyman’s article is motivated by fundamental questions regarding the ontological (relating to existence, reality) status of an organism—does an organism possess existence distinguishable from its molecular composition and social environment? His article reviews the role played by computational biology models in shedding light on these questions. This review draws on models from molecular kinetics to niche construction, all of which trace biological processes to a causal, and therefore existent, source.

I believe this issue of *Briefings in Bioinformatics* provides an interesting ‘snapshot’ of contemporary developments in bioinformatics and computational biology. It suggests that both areas are still evolving (pushed mainly by advances in experimental techniques) and possibly merging into a broader discipline. Moreover, the challenges these fields have to face in the future are likely to be even more complex.

Clearly, a single issue cannot claim to be complete or even comprehensive nor representative. I still believe that this issue is useful as a ‘yardstick’ to gauge ongoing and anticipate future developments in the fields of bioinformatics, computational biology and even systems biology.

Werner Dubitzky

References

1. Howe D, Costanzo M, Fey P, *et al.* Big data: the future of biocuration. *Nature* 2008;**455**:47–50.